2) Vanishing Gradients & Remedies

a) Vanishing Gradient Problem:
In RNNs, during Backpropagation Through Time (BPTT), gradients are repeatedly multiplied by small values at each time step, causing them to shrink exponentially.
As a result, **earlier layers receive almost no gradient updates**, making it difficult for the model to learn **long-range dependencies** in sequential data.

b) Architectural Solutions:

1. **LSTM (Long Short-Term Memory):**
   Uses gating mechanisms (input, forget, and output gates) and a cell state to **preserve and control information flow**, preventing gradients from vanishing over long sequences.

2. **GRU (Gated Recurrent Unit):**
   Simplifies LSTM with **update and reset gates** that regulate information flow and maintain longer-term dependencies by allowing gradients to pass through more directly.

c) Training Technique (Non-Architectural):
Gradient Clipping:
Caps the gradients to a fixed threshold during backpropagation, preventing them from becoming too small or too large, thus **stabilizing training and improving convergence**.